

## Galaxy and Mass Assembly: the evolution of bias in the radio source population to $z$ 1.5

Article (Published Version)

Lindsay, S N, Jarvis, M J, Santos, M G, Brown, M J I, Croom, S M, Driver, S P, Hopkins, A M, Liske, J, Loveday, J, Norberg, P and Robotham, A S G (2014) Galaxy and Mass Assembly: the evolution of bias in the radio source population to  $z$  1.5. *Monthly Notices of the Royal Astronomical Society*, 440 (2). pp. 1527-1541. ISSN 0035-8711

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/54684/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# Galaxy and Mass Assembly: the evolution of bias in the radio source population to $z \sim 1.5$

S. N. Lindsay,<sup>1★</sup> M. J. Jarvis,<sup>2,3</sup> M. G. Santos,<sup>3,4</sup> M. J. I. Brown,<sup>5</sup> S. M. Croom,<sup>6</sup>  
S. P. Driver,<sup>7,8</sup> A. M. Hopkins,<sup>9</sup> J. Liske,<sup>10</sup> J. Loveday,<sup>11</sup> P. Norberg<sup>12</sup>  
and A. S. G. Robotham<sup>7,8</sup>

<sup>1</sup>Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, Hatfield AL10 9AB, UK

<sup>2</sup>Oxford Astrophysics, Department of Physics, Keble Road, Oxford OX1 3RH, UK

<sup>3</sup>Physics Department, University of the Western Cape, Bellville 7535, South Africa

<sup>4</sup>CENTRA, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais 1, P-1049-001 Lisboa, Portugal

<sup>5</sup>School of Physics, Monash University, Clayton, VIC 3800, Australia

<sup>6</sup>Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW 2006, Australia

<sup>7</sup>International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, Crawley, WA 6009, Australia

<sup>8</sup>Scottish Universities' Physics Alliance (SUPA), School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK

<sup>9</sup>Australian Astronomical Observatory, PO Box 915, North Ryde, NSW 1670, Australia

<sup>10</sup>European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany

<sup>11</sup>Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK

<sup>12</sup>Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK

Accepted 2014 February 20. Received 2014 February 17; in original form 2013 January 14

## ABSTRACT

We present a large-scale clustering analysis of radio galaxies in the Very Large Array Faint Images of the Radio Sky at Twenty-cm survey over the Galaxy and Mass Assembly (GAMA) survey area, limited to  $S_{1.4\text{GHz}} > 1\text{ mJy}$  with spectroscopic and photometric redshift limits up to  $r < 19.8$  and  $< 22\text{ mag}$ , respectively. For the GAMA spectroscopic matches, we present the redshift space and projected correlation functions, the latter of which yielding a correlation length  $r_0 \sim 8.2 h^{-1}\text{ Mpc}$  and linear bias of  $\sim 1.9$  at  $z \sim 0.34$ . Furthermore, we use the angular two-point correlation function  $w(\theta)$  to determine spatial clustering properties at higher redshifts. We find  $r_0$  to increase from  $\sim 6$  to  $\sim 14 h^{-1}\text{ Mpc}$  between  $z = 0.3$  and  $1.55$ , with the corresponding bias increasing from  $\sim 2$  to  $\sim 10$  over the same range. Our results are consistent with the bias prescription implemented in the SKA Design Study simulations at low redshift, but exceed these predictions at  $z > 1$ . This is indicative of an increasing (rather than fixed) halo mass and/or active galactic nuclei fraction at higher redshifts or a larger typical halo mass for the more abundant Fanaroff and Riley Class I sources.

**Key words:** surveys – galaxies: active – large-scale structure of Universe – radio continuum: galaxies.

## 1 INTRODUCTION

In recent decades, measurements of the cosmic microwave background (CMB) radiation (e.g. Komatsu et al. 2011) have proven the very early Universe to be remarkably isotropic with tiny under- and overdensities that have grown to form the vast and intricate structures we see today. Galaxies and clusters observed today are far removed from their almost homogeneous beginnings and we require large numbers of them to piece together a statistical picture on cosmological scales. Clustering measures on large scales can be

used to investigate not only the relationships between galaxy populations found by various techniques probing different epochs and masses, but also broader cosmological phenomena such as baryon acoustic oscillations (BAO; e.g. Eisenstein et al. 2005; Percival et al. 2010; Blake et al. 2011), cosmic magnification (e.g. Scranton et al. 2005; Wang et al. 2011) or the integrated Sachs–Wolfe (ISW) effect (e.g. McEwen et al. 2007; Giannantonio et al. 2008; Raccanelli et al. 2008).

Cosmological applications require information about the gravitating mass distribution in the Universe, which in a  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology is strongly tied to the dark matter distribution. Direct observations tell us only about the baryonic matter, from which we must infer the dark matter distribution. Various tools

★E-mail: s.lindsay2@herts.ac.uk

exist for measuring the clustering signal of an observed source catalogue, such as nearest neighbour measures (e.g. Bahcall & Soneira 1983), counts-in-cells (e.g. Magliocchetti et al. 1999; Blake & Wall 2002b; Yang & Saslaw 2011), correlation functions (e.g. Groth & Peebles 1977; Bahcall & Soneira 1983; Blake & Wall 2002a,b; Croom et al. 2005) and power spectra (e.g. Cole et al. 2005; Percival et al. 2007; Komatsu et al. 2011). Because of its relative simplicity to calculate, and relation to its Fourier transform (the power spectrum), the two-point spatial correlation function has become a standard in quantifying cosmological structure. A means by which we can quantify the extent to which the observable and dark matter are tied using the correlation function is through the bias parameter  $b(z)$ . This relates the spatial correlation function (see Section 5) of an observed galaxy population to that of the underlying dark matter. The bias quantifies the difference in the clustering of the dark matter haloes acting solely under gravity and of galaxies inhabiting those haloes with other effects making their structure more or less diffuse. This has a heavy dependence on the galaxy masses and the epoch under consideration (e.g. Seljak & Warren 2004).

Local galaxies with masses comparable to the Milky Way are a relatively unbiased tracer of mass, with a present day correlation length (the clustering scale at which the correlation function falls below unity; see Section 5) of  $r_0 = 5.4 h^{-1}$  Mpc found in the early Center for Astrophysics (CfA) Redshift Survey (Davis & Peebles 1983). While the power-law slope is consistently found to be  $\sim 2$  in most studies, large variations in  $r_0$  are found depending on the populations and respective epochs being observed, from  $\sim 5 h^{-1}$  Mpc for local galaxies to  $\sim 25 h^{-1}$  Mpc for Abell clusters (e.g. Bahcall & Soneira 1983). At the lower end of the clustering scale, Saunders, Rowan-Robinson & Lawrence (1992) find  $r_0 \sim 4 h^{-1}$  Mpc for IRAS starburst galaxies, and using the 2dF QSO Redshift Survey, Croom et al. (2001) show a scale roughly constant with redshift at  $\sim 5 h^{-1}$  Mpc for quasars at  $0 \lesssim z \lesssim 2.5$  (see also Croom et al. 2005; Ross et al. 2009; Ivashchenko, Zhdanov & Tugay 2010). Likewise, Kovač et al. (2007) find that  $z \sim 4.5$  Ly $\alpha$  emitters exhibit a relatively short clustering length of  $4.6 h^{-1}$  Mpc, consistent with their being progenitors of Milky-Way-type local galaxies (see Nilsson & Meisenheimer 2009 and references therein), and note the similarity with Lyman-break galaxies (LBGs) at  $z \sim 3.8$  and  $\sim 4.9$ , suggesting that the two populations reside in the same host haloes but with a relatively low duty cycle. Local  $L \gtrsim L_*$  ellipticals, however, are a highly clustered population with various authors finding  $r_0 \sim 7\text{--}12 h^{-1}$  Mpc (e.g. Guzzo et al. 1997; Willmer, da Costa & Pellegrini 1998; Norberg et al. 2002). Similarly high clustering lengths are found for extremely red objects (EROs) at  $z \sim 1$  (e.g. Daddi et al. 2001; McCarthy et al. 2001; Roche et al. 2002) and distant red galaxies (DRGs) at  $z \sim 1\text{--}2$  (Grazian et al. 2006; Foucaud et al. 2007), both posited as progenitors of the local bright ellipticals. Indeed, similar results for radio galaxies (e.g. Cress et al. 1996; Overzier et al. 2003) give weight to the suggestion by Willott, Rawlings & Blundell (2001) that EROs and radio galaxies are identical and seen at different evolutionary stages, based on their findings of ERO-like hosts of radio galaxies in the 7C Redshift Survey.

Radio surveys are ideal for the purpose of carrying out such large-scale statistical measurements for a number of reasons. From a logistical point of view, radio wavelengths are useful because they occupy a uniquely broad atmospheric window in the electromagnetic spectrum, not being significantly absorbed by Earth's atmosphere. This allows observations to be carried out from terrestrial telescopes, reducing costs compared with launching and operating space telescopes in orbit. In a wide radio survey with relatively

shallow flux density limit ( $S_{1.4} \gtrsim 1$  mJy) extragalactic sources dominate in the form of synchrotron radiation in active galaxies with supermassive black holes at their centre. By observing such powerful sources, and with the long radio wavelengths being immune to absorption by dust, the interstellar medium (ISM) and our own atmosphere, ground-based telescopes are able to observe the whole sky relatively unobscured and to very high redshifts.

Of particular significance for the science discussed here is that these radio-loud active galactic nuclei (AGN) predominantly reside in massive elliptical galaxies (e.g. Jarvis et al. 2001; Willott et al. 2003; McLure et al. 2004; Herbert et al. 2011) and, as such, provide good tracers of large-scale clustering and the underlying dark matter distribution. Radio surveys have been exploited with this aim for some time, such as the 4.85 GHz 87GB (Kooiman, Burns & Klypin 1995) and Parkes–MIT–NRAO (Loan, Wall & Lahav 1997) surveys, the 325 MHz Westerbork Northern Sky Survey (WENSS; Rengelink et al. 1998), 1.4 GHz Very Large Array (VLA) surveys, Faint Images of the Radio Sky at Twenty-cm (FIRST; Becker, White & Helfand 1995) and the NRAO VLA Sky Survey (NVSS; Condon et al. 1998; e.g. Cress et al. 1996; Blake & Wall 2002a; Overzier et al. 2003; Best et al. 2005; Fine et al. 2011), but their potential uses are limited without additional data at other wavelengths. While the large redshift range of radio surveys can be advantageous over more targeted optical surveys, this can also wash out the clustering signal due to the superposition of the influence of differently behaving subpopulations (e.g. Wilman et al. 2008). This effect can be alleviated by the use of complementary optical surveys to aid the separation of radio sources by simple redshift binning or other methods.

As technology advances, observations at radio wavelengths will be able to constrain cosmological parameters with more precision than ever (Camera et al. 2012; Raccanelli et al. 2012), as well as provide the data for much more galactic and extragalactic science (Norris et al. 2011). Moving forward, the large numbers of sources observed in various overlapping surveys will make multiwavelength studies not only more common, but essential to reliably identify objects by their emission across the entire spectrum and delineate the clustering properties of various subpopulations of any given individual survey.

It is with this goal in mind that in this paper we combine the available radio data from the FIRST survey, and optical/near-infrared photometry and spectra from Galaxy and Mass Assembly (GAMA; Driver et al. 2011), Sloan Digital Sky Survey (SDSS; York et al. 2000) and UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007) to measure the large-scale clustering properties of the  $S_{1.4} > 1$  mJy radio emitting galaxies. We choose to use FIRST over NVSS for its marginally greater depth and considerably better resolution which is essential for meaningful cross-identification with the optical surveys. Furthermore, we employ a parent redshift distribution from the simulated radio catalogue of the SKA Design Study (SKADS; Wilman et al. 2008) which has the express intention of acting as a test bed to inform the use of the forthcoming Square Kilometre Array (SKA).

The outline of this paper is as follows. Section 2 describes the surveys and catalogues used, while Sections 3 and 4 detail the angular correlation function measurements and their deprojection to infer spatial clustering properties. Section 5 describes direct measurements of the spatial clustering properties through the spatial correlation function and projected correlation function, and Section 6 describes the linear bias found from each of these methods out to  $z \sim 1.5$ . In Section 7 we summarize our results and conclusions.

The cosmological model used throughout this paper is the flat,  $\Lambda$ CDM concordance cosmology where  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  and  $\sigma_8 = 0.8$ . All distances are kept in units of  $h^{-1}$  Mpc where  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $h$  is not explicitly assumed.

## 2 DATA

To infer the spatial clustering parameters of the radio galaxy population, redshifts of the sources (or their redshift distribution, at least) are required. We use spectroscopic redshifts from the GAMA survey, supplemented by photometric redshifts calculated using optical and near-infrared photometry from SDSS and UKIDSS (described in Smith et al. 2011). These redshift catalogues are cross-matched with radio sources from the FIRST survey to assign optical counterparts and redshifts to each radio source.

### 2.1 Radio surveys

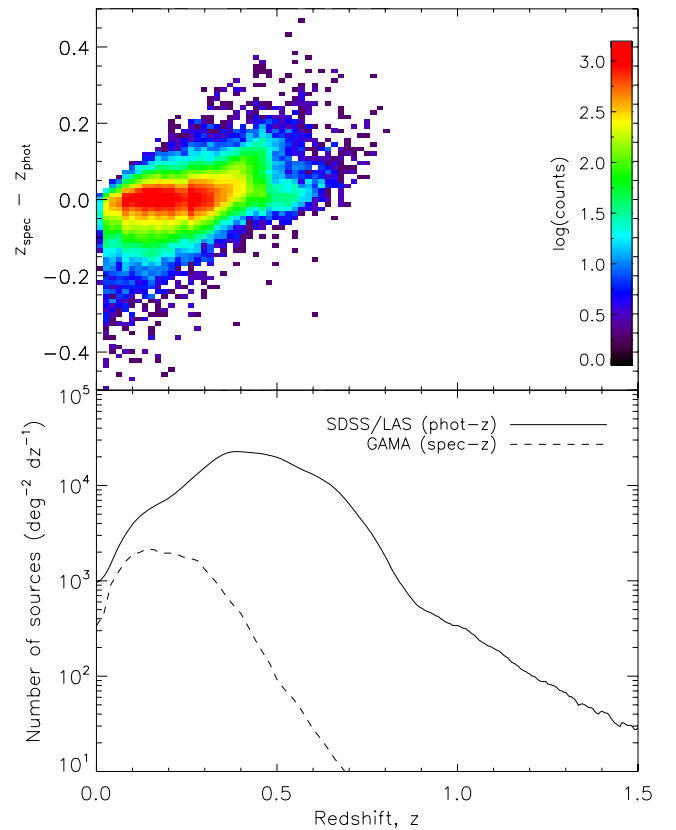
The FIRST survey (Becker et al. 1995) was carried out at an observing frequency of 1.4 GHz with the VLA in B configuration. The most recent catalogue<sup>1</sup> contains 946 464 sources covering over 10 000  $\text{deg}^2$  ( $\sim 8500$  in the northern galactic cap and  $\sim 1500$  in the south) with an angular resolution of 5.4 arcsec [full width at half-maximum (FWHM)] to a completeness of 95 per cent at  $S_{1.4\text{GHz}} > 2 \text{ mJy}$ . A number of sidelobes are spuriously counted in the raw catalogue and an oblique decision-tree program developed by the FIRST survey team (White et al. 1997) finds probabilities of each catalogue entry representing extended activity from a nearby bright source. Those entries with a sidelobe probability of  $> 0.1$  have been excluded for the purposes of this analysis ( $< 20$  per cent of the catalogue), leaving 723 934 sources above 1 mJy.

### 2.2 Optical and near-infrared surveys

The GAMA (Driver et al. 2011) survey has been in operation since 2008, using the 3.9-m Anglo-Australian Telescope (AAT) and the AAOmega spectrograph to build a  $> 98$  per cent complete redshift catalogue of  $\sim 140\,000$  galaxies to a depth of  $r < 19.4$  or 19.8. Two gratings with central wavelengths of 4800 and 7250  $\text{\AA}$  are used, giving continuous wavelength coverage of the range 3720–8850  $\text{\AA}$  with resolution  $\sim 3.5 \text{ \AA}$  (in the blue channel) and  $\sim 5.5 \text{ \AA}$  (in the red channel). Spectroscopic redshifts are found in real time using the RUNZ code developed for the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001) and given a quality flag  $Q$  of 0–4, where  $Q \geq 3$  is a probable or certain redshift worthy of publication (Hopkins et al. 2013). Corresponding photometry is also available in UKIDSS and SDSS bands (Hill et al. 2011).

GAMA is divided over three  $12^\circ \times 4^\circ$  equatorial regions centred at  $\alpha = 135^\circ$  (9h),  $180^\circ$  (12h) and  $217.5^\circ$  (15h), covering 144  $\text{deg}^2$ . These were chosen to be large enough in area to fully sample  $\sim 100 h^{-1} \text{ Mpc}$  structures at  $z \sim 0.2$ , and for their overlap with existing surveys. The data used for this paper come from the first 3 yr of observations (GAMA Phase I; Driver et al. 2011).

The overlap of the GAMA survey area with other surveys allows for the collection of sufficient photometry to be able to generate a large contribution to the redshift catalogue using photometric redshifts. The SDSS has  $\sim 10\,000 \text{ deg}^2$  of sky coverage in five bands (*ugriz*), with the southernmost stripes covering the three GAMA



**Figure 1.** Top: comparison of SDSS/LAS photometric redshift estimates with GAMA spectroscopic ( $Q \geq 3$ ) redshifts. Bottom: redshift distributions of the full photometric and spectroscopic catalogues (galaxies only).

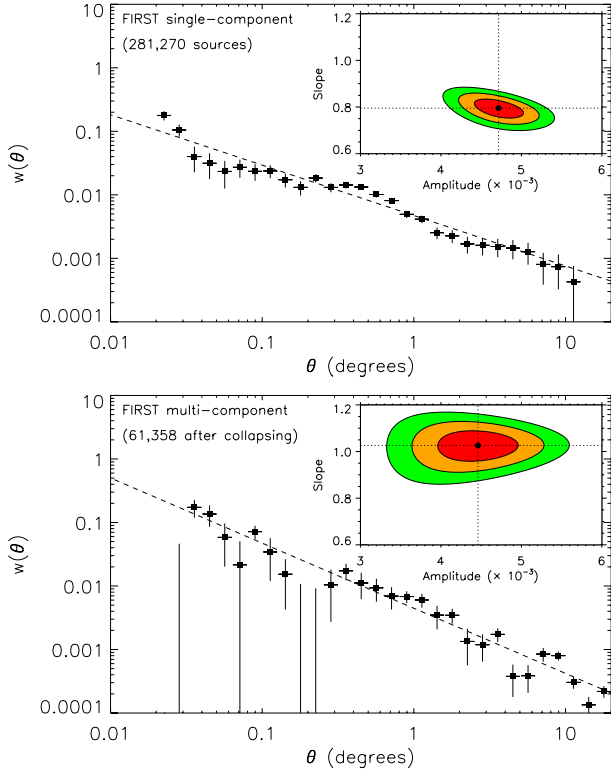
fields. The UKIDSS uses the 3.8-m United Kingdom Infrared Telescope (UKIRT) consisting of several subsurveys, including the Large Area Survey (LAS). UKIDSS LAS complements the SDSS, covering over 2000  $\text{deg}^2$  in the *YJHK* bands within the SDSS regions and overlapping with GAMA. We therefore have photometry in up to nine optical and infrared bands with which to obtain photometric redshifts for those galaxies without spectroscopy (further details are given in Smith et al. 2011). For those sources detected in the SDSS imaging data but without GAMA spectroscopic redshifts, we impose a magnitude limit of  $r < 22$ , i.e. within the limit of the SDSS imaging data. Fig. 1 shows a comparison between spectroscopic and photometric redshifts, where we find an rms difference of 0.07, and their distributions.

### 2.3 Collapsing of multicomponent radio sources

In order to address the inevitable issue of extended radio sources resulting in multiple detections for one host galaxy, of which perhaps none corresponds to the core itself (and therefore any associated optical source), we have followed Cress et al. (1996) in applying a collapsing radius of 72 arcsec ( $0^{\circ}02'$ ) to the FIRST catalogue. Any FIRST sources within this radius of one another are grouped and combined to form a single entry positioned at the flux-weighted average coordinates of the group and attributed with their summed flux density. This precludes us probing the correlation function to angular scales smaller than 72 arcsec, but ensures that only extremely extended sources will be mistakenly treated as independent galaxies. The procedure does, however, suffer from introducing far

<sup>1</sup> [http://sundog.stsci.edu/first/catalogs/readme\\_12feb16.html](http://sundog.stsci.edu/first/catalogs/readme_12feb16.html)





**Figure 2.** Angular correlation functions for a  $115^\circ \times 64^\circ$  area of the FIRST survey at 1 mJy, with 68, 90 and 95 per cent contours for the power-law parameter fits. The top panel corresponds to those FIRST sources which have no neighbours within the 72 arcsec collapsing radius adopted. The bottom panel corresponds to the remaining (assumed) extended sources post-collapsing.

greater positional uncertainty to the assumed core position given the inherent variation in morphology in the observed objects. Asymmetric multiple-component galaxies or independent galaxies with small angular separations by chance, in particular, will interfere with the reliability of the collapsed catalogue. From the original FIRST catalogue, 227 606 sources are collapsed in 98 225 distinct groups, leaving a  $>1$  mJy catalogue of 585 473 sources.

Returning to the option of supplementing FIRST with NVSS data to better deal with extended sources, when matching collapsed FIRST sources or NVSS sources directly to the optical catalogues, we find no evidence of a significant difference between the results.

Fig. 2 shows the angular correlation function (discussed in Section 3) of a large area ( $115^\circ \times 64^\circ$ ) of the FIRST survey, separately for single sources and collapsed extended sources. The single sources demonstrate a well constrained power law form while the collapsed sources give more erratic results, beyond the simple Poisson errors due to their smaller numbers. However, the results are still broadly compatible and the signal of the combined, collapsed catalogue is dominated by the well-behaved single source power law. The two samples would not be expected to have the same  $w(\theta)$ , however, as the collapsed sources are dominated by extended Fanaroff and Riley Class I (FRI) and Fanaroff and Riley Class II (FRII) AGN, in contrast to the largely compact single-component sources, which are a combination of compact AGN and low-redshift star-forming galaxies.

**Table 1.** Spatial boundaries and surface densities ( $\bar{\sigma}$ ) for the three GAMA/SDSS/LAS fields used (comprising both spectroscopic and photometric redshifts). Because of incomplete coverage in one corner of the 9h field, the area is  $\sim 0.4 \text{ deg}^2$  smaller in area than the full rectangular 12h and 15h fields. All subsequent analysis accounts for this.

Field	RA range ( $^\circ$ )	Dec. range ( $^\circ$ )	Optical sources	$\bar{\sigma}$ ( $\text{deg}^{-2}$ )
9h	[128.0, 142.0]	[−2.0, 3.0]	633 229	9101
12h	[172.5, 186.5]	[−2.9, 2.1]	651 148	9303
15h	[210.5, 224.5]	[−2.0, 3.0]	653 417	9336

## 2.4 Optical identification of FIRST sources

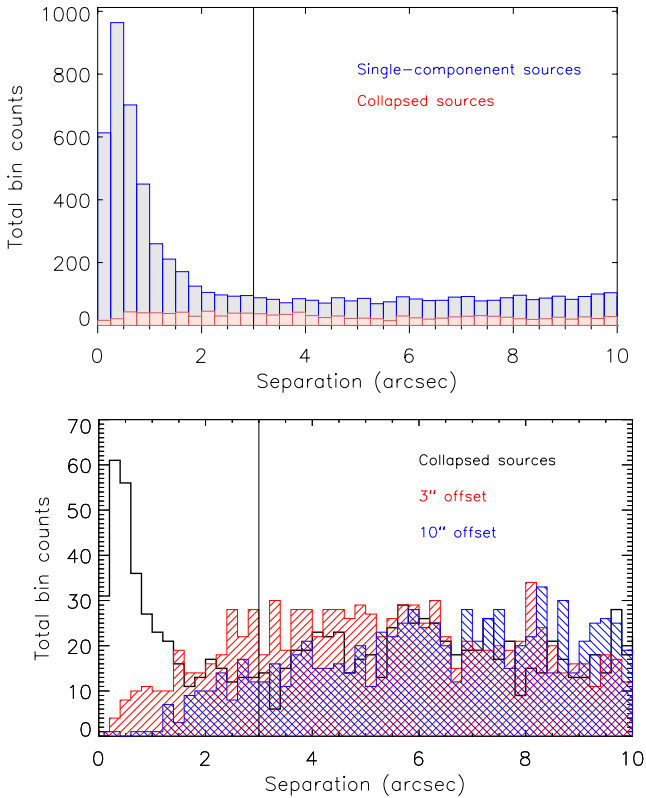
An optical catalogue somewhat larger than the GAMA survey area is used for the purpose of cross-matching with the radio sources (see Table 1), making use of photometric redshifts determined using up to nine SDSS/UKIDSS LAS bands (*ugrizYJHK*). This allows for greater source counts and improved statistics for the analysis, as well as expanding to a region completely containing the GAMA fields as well as other surveys like *Herschel*-Astrophysical Terahertz Large Area Survey (ATLAS; Eales et al. 2010). The expanded region comprises three  $14^\circ \times 5^\circ$  fields expanding from the three  $12^\circ \times 4^\circ$  GAMA fields.

Of the  $\sim 6$  million optical/IR sources in this area with at least a photometric redshift, approximately half are removed on the basis of the star–galaxy separation technique by Baldry et al. (2010), leaving only the likely extragalactic sources to be matched to the radio. This separation is based on  $J - K$  and  $g - i$  colours rather than simply removing point-like sources and therefore fewer quasars are mistakenly discarded, provided they lie away from the stellar locus. As a second measure, a minimum redshift criterion of  $z > 0.002$  was added to filter out the nearby objects (some with significantly negative redshifts) which are assumed to be stellar in origin or to correspond to highly extended extragalactic sources, although we note that this may also filter out a small number of real low-redshift extragalactic radio sources.

For the purpose of radio–optical cross-matching, given the positional accuracies of the catalogues ( $<1$  arcsec for FIRST,  $\sim 0.1$  arcsec for SDSS) a simple nearest-neighbour match can be reliable. Sullivan et al. (2004) find this method to produce very similar catalogues to the likelihood ratio protocol of Sutherland & Saunders (1992) in their work. The likelihood ratio technique is often used to identify radio sources (Cileigi et al. 2005; Gonzalez-Solares et al. 2005; Afonso et al. 2006; McAlpine et al. 2012) but here we instead use the simpler method tested by Sullivan et al. (2004) and adopted by e.g. El Boucheffry & Cress (2007). The 13 346 radio sources within the three fields were all found to have optical counterparts within 1 arcmin. By inspection of the separation distribution shown in Fig. 3, any separations of more than a few arcseconds are likely to be random. A compromise must be reached between including the greatest number of our radio sample and ensuring reliable optical cross-identifications. The expected contamination of the remaining objects by chance proximity of optical sources can be expressed through the simple equation:

$$P_c = \pi r_s^2 \sigma_{\text{opt}}, \quad (1)$$

where  $r_s$  is the search radius cut-off and  $\sigma_{\text{opt}} = 9.23 \times 10^3 \text{ deg}^{-2}$ , the surface density of the optical/IR galaxy catalogue. This gives a 0.9 per cent contamination rate for a 2 arcsec cut-off and 2.0 per cent for a 3 arcsec cut-off. We place our separation limit at 3 arcsec, below which the distribution in Fig. 3 becomes visibly more dense. This leaves 3886 (29 per cent) of the original FIRST radio sources,



**Figure 3.** Top: distribution of offsets between FIRST sources and the nearest GAMA/SDSS source, with single-component sources (blue) placed on top of the collapsed sources (red). The vertical solid line indicates the proposed separation cut-off of 3 arcsec to define a good radio–optical cross-match, below which counts rise above the background level. Collapsed sources account for  $\sim 10$  per cent of matches within the cut-off. Bottom: distribution of separations for the collapsed radio sources (black). The filled histograms show the result of assigning a random Gaussian offset of  $\sigma = 3$  arcsec (red) and 10 arcsec (blue) before the cross-matching showing a strong decline in matches within 2 arcsec.

of which 422 correspond to collapsed multicomponent sources, an estimated 78 are spurious and 1424 (42 per cent) have good quality GAMA spectroscopic redshifts.

While positional coincidence is appropriate for defining positive identifications of single isolated radio sources, it is not necessarily so for the collapsed multiple sources, as this introduces uncertainty into the assumed position of the optical core of the host galaxy. These sources are in the minority, but to assess the reliability of their matches, we repeat the cross-matching process with the radio positions randomly displaced by 3 and 10 arcsec Gaussian distributions. Fig. 3 (bottom) shows the resulting separation histograms. While all three iterations of the procedure agree above the 3 arcsec matching cut-off, the significant peak at low separations is completely lost after shifting the radio positions. There will almost certainly be a minority of collapsed sources for which we do not accurately find a radio core position, but this test shows that for those where the collapsing is successful, reliable optical matches are being found.

## 2.5 Redshift distributions

In order to study the spatial clustering properties of our galaxy samples, we require knowledge of their redshifts, either individually

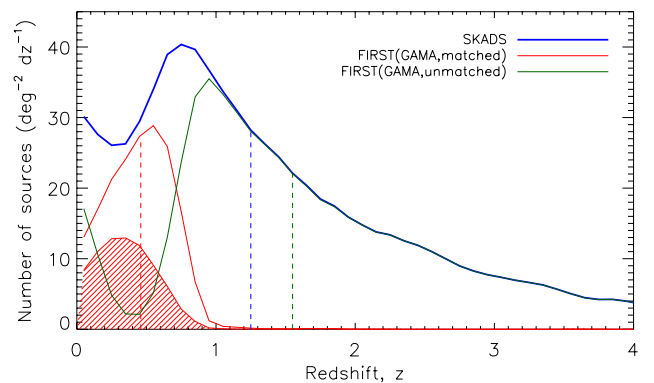
**Table 2.** Areas and surface densities of  $S_{1.4} > 1$  mJy radio sources in the FIRST catalogues, the GAMA counterparts to the radio sources and also for the SKADS simulated data set. Note that the low density of sources in the whole (post-collapse) FIRST catalogue compared to the prediction from SKADS may be explained by the incompleteness at  $< 2$  mJy (Sections 2.6 and 4.2.1) and the effects of resolution bias (Section 2.5).

Catalogue	No. of sources	Area (deg <sup>2</sup> )	$\bar{\sigma}$ (deg <sup>-2</sup> )
FIRST	585 473	$\sim 10\,000$	$\sim 58.5$
GAMA	3886	210	18.5
SKADS	32 061	398	80.6

or at least their distribution. We have spectroscopic or photometric redshifts for those FIRST radio sources with optical identifications, but do not have direct information for the optically unidentified radio sources. To estimate the complete redshift distribution as a function of radio flux density limit, we use the SKADS simulation which is based on a range of observed luminosity functions (see Wilman et al. 2008 for full details). Table 2 compares the area and source density of the SKADS simulated catalogue with our large FIRST sample, and the GAMA cross-matched sample.

Fig. 4 shows the redshift distribution of the SKADS catalogue at 1 mJy and the cross-matched subset of FIRST. By making the assumption that FIRST sources (the wider catalogue as well as just those within the GAMA fields) should have a similar distribution to that from SKADS, we may compare directly the clustering of the real and simulated catalogues. Furthermore, as shown in Fig. 4, a third distribution can be inferred by subtracting the cross-matched distribution from the SKADS ‘parent’ distribution (following similar work by Passmoor et al. 2013). This describes the remaining unmatched radio sources with no direct redshift measurements.

It is clear that the distribution of unmatched sources shows an up-turn towards low redshift ( $z < 0.2$ ), where there appears to be a large fraction of sources in the SKADS simulation that are either not detected in FIRST or are in FIRST but do not have a counterpart in GAMA. The latter explanation of this is unlikely, as we would expect that the majority of relatively bright radio sources at  $z < 0.2$



**Figure 4.** Redshift distributions of the 1 mJy SKADS (blue) and GAMA catalogues (red), where the SKADS  $N(z)$  is assumed to be the redshift distribution of FIRST sources. The distribution of the radio sources not identified in GAMA (green) is inferred assuming a SKADS-like parent distribution and subtracting the GAMA distribution. Dotted lines mark the median redshifts for each set of objects and the filled distribution describes the GAMA spectroscopic redshifts (i.e. removing SDSS/LAS photometric redshifts).

to have a reasonably bright optical counterpart. We therefore suggest that these sources, which are dominated by star-forming galaxies in the SKADS simulation and are predicted to be there based on low-redshift far-infrared luminosity functions (see Wilman et al. 2010), are likely to be resolved out by the VLA observations in the B-Array configuration used for the FIRST survey (see also Jarvis et al. 2010).

Simpson et al. (2012) also noted a similar deficit of low-redshift sources compared with SKADS predictions in their radio survey of the Subaru/*XMM-Newton* Deep field, and attributed this to the effects of resolution bias. This resolution bias refers to the number of faint resolved sources missing from a peak-flux-density-limited survey because their extended emission is not detected by the radio interferometer. Bondi et al. (2003) performed detailed simulations to determine the effect of this bias on their deep VLA catalogue and their estimates suggest that around 25 per cent of sources could be missing due to this resolution bias, and that these would be concentrated at low redshift. Any low-redshift star-forming galaxies that have radio flux density detectable in FIRST are unlikely to fall below the SDSS depth ( $r < 22$ ), although low surface brightness optical incompleteness may play a role if the radio data were deeper.

Thus we suggest that the discrepancy between the simulated redshift distribution and the observed redshift distribution at low redshift is consistent with being due to resolution bias. We also note that some of the sources in our cross-matched sample could also be removed by our removal of sources with photometric redshifts  $z < 0.002$ .

## 2.6 Completeness

While the FIRST radio catalogue contains a large number of  $\sim 1$  mJy sources, it is incomplete below 2–3 mJy. Discarding these fainter radio sources would significantly affect the size of our samples and the statistical significance of our results, but the effects of the incompleteness at the mJy level must be corrected for. To account for the noise variations across the survey area we use an rms noise map of the FIRST survey and their source detection criterion of  $S - 0.25 \text{ (mJy)} > 5 \times \text{rms}$  to determine whether a radio source randomly drawn from the SKADS simulated catalogues would be detected in the FIRST data over the GAMA fields. We use only those sources from SKADS with  $S_{1.4} > 0.5$  mJy, and if the SKADS radio source fulfils the FIRST survey detection criteria then that source is retained, whereas if the source falls below the detection threshold then the source is omitted. This process not only provides an estimate of the completeness of the survey but also allows us to compile a catalogue of sources, with both flux density and redshift information, that is subsequently used as our random source catalogue for calculating the angular and spatial correlation function (see Sections 3–5). This ensures that we fully account for any noise variation and incompleteness across the survey area. The model redshift distribution of radio sources in our survey regions is thus modified according to these noise variations and it is this modified redshift distribution which is used throughout in determining the spatial correlation function.

It is unlikely that this incompleteness is responsible in itself for the dearth of low-redshift radio matches (cf. Section 2.5), which make up a minority of the sources at  $\sim 1$  mJy. One would expect to preferentially lose the higher redshift AGN at around the flux density limit. For example, the SKADS simulation features  $\sim 5$  low-redshift star-forming galaxies per square degree compared to  $\sim 30$  AGN at higher redshifts, in the flux density range of 1–2 mJy.

**Table 3.** A summary of the samples for which the clustering properties will be measured: GAMA/SDSS cross-matched radio sources (further split into  $z > 0.5$  and  $z < 0.5$  samples), a large sample of FIRST sources, a smaller sample of all FIRST sources within the GAMA fields and the remaining FIRST sources in these fields after removing GAMA cross-matches (all post-collapse of multiple sources). Redshift distributions are shown in Fig. 4, and the enlarged GAMA area footprint is shown in Fig. 5.

Sample	Area (deg <sup>2</sup> )	No. of sources	Redshift	
			Limit	$z_{\text{med}}$
GAMA(matched)	210	3886	–	0.48
GAMA(matched)	210	2156	$z < 0.5$	0.30
GAMA(matched)	210	1730	$z > 0.5$	0.65
FIRST	5922	342 615	–	1.21
FIRST(G)	210	13 346	–	1.21
FIRST(G-unmatched)	210	9460	–	1.55

## 2.7 Sample catalogues

Table 3 lists the various samples to be investigated, comprising one large FIRST sample and five subsamples of FIRST sources within the GAMA fields. The large FIRST sample is over a single  $115^\circ \times 64^\circ$  patch ( $127.5^\circ < \alpha < 242.5^\circ$ ;  $0^\circ < \delta < 64^\circ$ ), while two smaller subsets contain all FIRST sources found within the enlarged GAMA footprint defined in Table 1 (FIRST(G)) and only those which are unmatched with GAMA sources (FIRST(G-unmatched)), shown in Fig. 5. The matched GAMA sources are also subsequently split into low- and high-redshift subsets in order to attempt to demonstrate any redshift evolution. The optical cross-matching preferentially identifies low-redshift radio sources, allowing us to use the unmatched sources to probe higher redshifts ( $z_{\text{med}} = 1.55$ ) than the matched or full samples, further helping to observe any redshift dependence of their clustering.

## 3 ANGULAR CORRELATION FUNCTION

A simple measure of the clustering of sources in an angular distribution is found in the angular two-point correlation function,  $w(\theta)$ . It is defined as the excess probability of finding a galaxy at an angular distance  $\theta$  from another galaxy, as compared with a Poissonian (unclustered) distribution (Peebles 1980):

$$\delta P = \sigma[1 + w(\theta)]\delta\Omega, \quad (2)$$

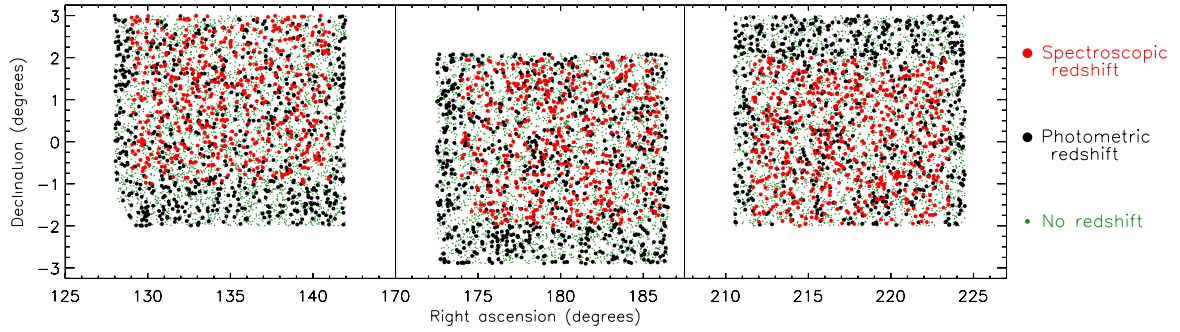
where  $\delta P$  is the probability,  $\sigma$  is the mean surface density and  $\delta\Omega$  is the surface area element.

A simple estimator defines  $w(\theta) = DD/RR - 1$ , where  $DD(\theta)$  and  $RR(\theta)$  represent the number of galaxy pairs separated by  $\theta$  in the real data and a random catalogue, respectively. Introducing an additional count of the cross-pair separations  $DR(\theta)$  allows for reduced variance, such as is found for our chosen estimator by Landy & Szalay (1993):

$$w(\theta) = \frac{n_r(n_r - 1)}{n_d(n_d - 1)} \frac{DD}{RR} - \frac{(n_r - 1)}{2n_d} \frac{DR}{RR} + 1, \quad (3)$$

where  $n_d$  and  $n_r$  are the number of real sources and random sources.

By averaging over several random data sets and using  $\overline{DR}$  and  $\overline{RR}$  or by using a more densely populated random catalogue, we may assume the statistical error in the random sets to be negligible. The random catalogues themselves have been autocorrelated



**Figure 5.** The angular distribution of the radio samples used. Optically identified FIRST sources are plotted in red (GAMA spectroscopic redshifts) or black (SDSS/UKIDSS photometric redshifts), and those without optical matches are plotted in green.

finding no significant deviation from zero even at extremes of angular separation where bin counts are lowest. The uncertainty on  $w$ , therefore, is often given by the Poisson error due to the  $DD$  counts alone:

$$\Delta w = \frac{1 + w(\theta)}{\sqrt{DD}}. \quad (4)$$

However, the errors in the correlation function depend on the  $DD$  counts beyond simple Poisson variance; adjacent bins are correlated, with each object contributing to counts across a range of separation bins. The errors are therefore calculated somewhat more rigorously using a bootstrap resampling technique (Ling, Barrow & Frenk 1986) whereby several data catalogues are constructed by randomly sampling (with replacement) the original set of objects. As such, in any given set, some sources are counted twice or more and some not at all. The resulting binned  $DD$  counts should give a mean approximately equal to the original data but allow us to calculate a variance for each bin, and therefore  $w(\theta)$  values. Cress et al. (1996), for example, found errors in  $w(\theta)$  for the early FIRST survey with a bootstrap resampling method. They found the Poisson error estimates to be too small by a factor of 2 on small scales ( $\sim 3$  arcmin) and more than an order of magnitude for larger scales ( $\sim 5^\circ$ ). We find Poisson error estimates to be consistently a factor of 2 smaller than the bootstrap error up to  $\sim 1^\circ$ , above which the ratio increases approximately exponentially.

The restricted survey area from which we can measure  $w(\theta)$  also results in a negative offset in the observed correlation function, known as the *integral constraint*. Expressed mathematically, the relation between the observed correlation function  $w_{\text{obs}}(\theta)$  and the genuine function  $w(\theta)$  is

$$w_{\text{obs}}(\theta) = w(\theta) - \sigma^2, \quad (5)$$

where  $\sigma^2$  represents the integral constraint (Groth & Peebles 1977) which can be approximated, following Roche & Eales (1999), by

$$\sigma^2 = \frac{\sum RR(\theta)w(\theta)}{\sum RR(\theta)}. \quad (6)$$

Traditionally,  $w(\theta)$  has been fitted by a power law (e.g. Peebles 1980) with a slope of  $\sim 0.8$  commonly found for the clustering of objects of various masses (Bahcall & Soneira 1983). While radio sources have been found to fit a distinct double power law (cf. Blake & Wall 2002a; Overzier et al. 2003), since we have collapsed the multiple component sources,  $w(\theta)$  should reduce to the canonical single power law form for our data.

We fit  $w(\theta)$  with a single power-law function of the form  $w(\theta) = A\theta^{1-\gamma}$ , fitted over the range  $0.02 < \theta < 10^\circ$ . This is done using the Metropolis–Hastings algorithm to obtain a Markov chain Monte Carlo (MCMC) simulation of  $10^6$  data points in  $\gamma$ – $\log(A)$  space. Parameter values quoted are minimum  $\chi^2$  values and  $1\sigma$  errors correspond to the region containing 68.3 per cent of the MCMC points. A Levenberg–Marquardt  $\chi^2$  minimization routine yields the same best-fitting values, and ordinary  $\chi^2$  contours coincide closely with those from the MCMC simulations.

Fig. 6 shows the results of this angular correlation function method for the five samples of FIRST radio sources within the GAMA regions, and we see clearly in the likelihood contours that the errors are smaller for the more numerous unmatched sources than the matched sources. The best-fitting parameters are shown in Table 4. The unmatched sample has a lower amplitude than the matched sample, but given these will come to represent complementary high- and low-redshift measurements, respectively, similar amplitudes do not imply similar clustering scales due to the increased angular diameter distance at high redshift.

#### 4 LIMBER INVERSION

The three-dimensional analogue of  $w(\theta)$  is the spatial two-point correlation function  $\xi(r)$ , which measures the excess probability, due to clustering, of finding a pair of objects separated by  $r \rightarrow r + \delta r$  as compared with a Poissonian (unclustered) distribution, defined as

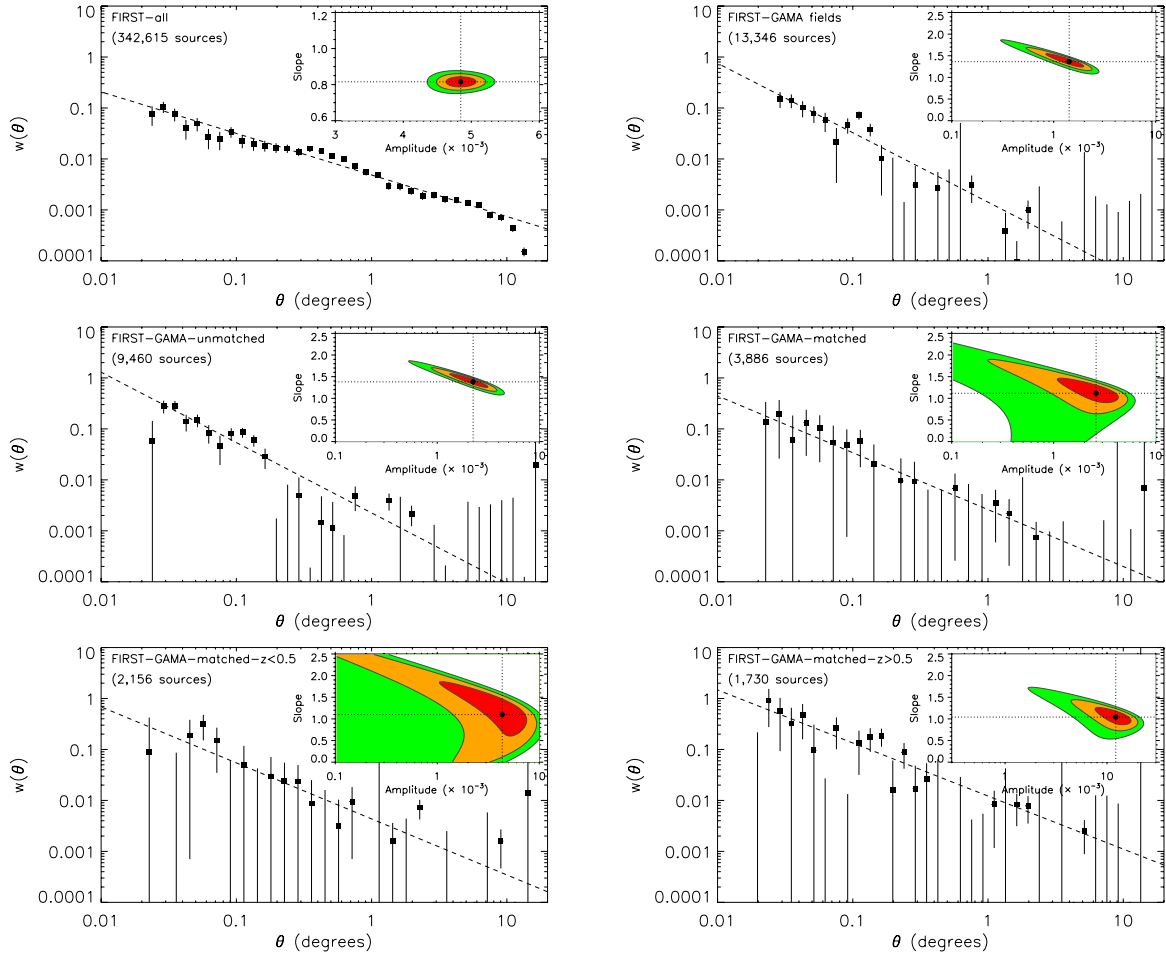
$$\delta P = n[1 + \xi(r)]\delta V, \quad (7)$$

where  $n$  is the mean number density of objects and  $\delta V$  the volume element.

If the redshift distribution of a set of objects is known, one may deproject the angular correlation function into the spatial correlation function. This is the purpose of the cosmological Limber equation (Limber 1953; Peebles 1980) for estimating the spatial correlation length,  $r_0$  (discussed further in Section 5). This is often more useful than computing the spatial correlation function,  $\xi(r)$ , directly as a complete set of individual redshifts is rarely available for a given survey, thus requiring the redshift distribution to be estimated via the luminosity function in order to deproject  $w(\theta)$ . Using the redshifts available for the objects in our catalogue, however, means we may apply a distribution directly from the data for the radio sources with optical counterparts.

An epoch-dependent form of the spatial correlation function is assumed (see e.g. de Zotti et al. 1990; Overzier et al. 2003, and





**Figure 6.** The angular correlation function for the five samples (sample name is denoted at the top left-hand corner of each panel) within the extended GAMA fields (with bootstrap resampling errors) and the larger FIRST sample in the bottom right-hand panel. The dashed lines show the best-fitting power law (over the range  $0.02 < \theta < 10^\circ$ ) and the inset contour plots show  $\chi^2$  parameter fits at 68, 90 and 95 per cent confidence levels.

**Table 4.** Clustering parameters measured covering real and simulated radio sources ( $S_{1.4} > 1$  mJy) with real (GAMA) and/or simulated redshift distributions (FIRST). Errors are  $1\sigma$  from  $10^6$  ( $A$ ,  $\gamma$ ) and  $10^5$  ( $r_0$ ,  $b$ ) MCMC calculations.

Sample	$N_{\text{obj}}$	$A (\times 10^{-3})$	$\gamma$	$z_{\text{med}}$	$r_0 (h^{-1} \text{ Mpc})$		$b (z = z_{\text{med}})$	
					$\epsilon = \gamma - 1$	$\epsilon = \gamma - 3$	$\epsilon = \gamma - 1$	$\epsilon = \gamma - 3$
FIRST-all	342 615	$4.87^{+0.12}_{-0.17}$	$1.82^{+0.02}_{-0.02}$	1.21	$11.82^{+0.43}_{-0.46}$	$8.20^{+0.41}_{-0.42}$	$4.39^{+0.19}_{-0.19}$	$3.14^{+0.16}_{-0.17}$
FIRST-GAMA fields	13 346	$1.31^{+0.35}_{-0.34}$	$2.35^{+0.11}_{-0.09}$	1.21	$11.27^{+0.82}_{-0.87}$	$10.53^{+0.64}_{-0.68}$	$6.40^{+0.42}_{-0.45}$	$5.95^{+0.48}_{-0.53}$
FIRST-GAMA-unmatched	9460	$2.21^{+0.62}_{-0.57}$	$2.39^{+0.11}_{-0.10}$	1.55	$14.41^{+1.23}_{-1.46}$	$13.60^{+0.83}_{-1.07}$	$10.06^{+0.49}_{-0.50}$	$9.45^{+0.58}_{-0.67}$
FIRST-GAMA-matched	3886	$2.76^{+0.94}_{-1.11}$	$2.15^{+0.12}_{-0.24}$	0.48	$8.24^{+1.75}_{-2.36}$	$6.72^{+1.81}_{-2.17}$	$2.65^{+0.98}_{-0.90}$	$2.13^{+0.90}_{-0.76}$
FIRST-GAMA-matched- $z < 0.5$	2156	$4.28^{+1.99}_{-1.89}$	$2.28^{+0.24}_{-0.51}$	0.30	$6.21^{+2.34}_{-4.34}$	$5.39^{+2.53}_{-3.92}$	$1.77^{+1.62}_{-1.23}$	$1.52^{+1.57}_{-1.07}$
FIRST-GAMA-matched- $z > 0.5$	1730	$12.26^{+2.90}_{-3.04}$	$2.04^{+0.12}_{-0.12}$	0.65	$17.07^{+4.50}_{-4.70}$	$10.67^{+3.22}_{-3.18}$	$5.74^{+2.56}_{-1.97}$	$3.56^{+1.59}_{-1.23}$

references therein):

$$\xi(r, z) = \left(\frac{r_0}{r}\right)^\gamma (1+z)^{\gamma-(3+\epsilon)}, \quad (8)$$

where  $r$  is in comoving units and  $\epsilon$  parametrizes the clustering model being assumed. Overzier et al. (2003) and Kim et al. (2011) offer three main models: *stable clustering* (where clusters have fixed physical size;  $\epsilon = 0$ ), *comoving clustering* (where clusters have fixed comoving size;  $\epsilon = \gamma - 3$ ) or *linear clustering* (growth under linear perturbation theory;  $\epsilon = \gamma - 1$ ). For a typical slope

found in the literature of  $\gamma \sim 2$ , these clustering models are stable, decaying and growing, respectively. Other authors, such as Elyiv et al. (2012), also apply  $\epsilon = -3$ . Given that  $\epsilon > 0$  (or  $< 0$ ) implies growing (or decaying) clustering, this final  $\epsilon$  value implies a more rapid clustering decay than the other models. In this paper we adopt  $\epsilon = \gamma - 3$  and  $\epsilon = \gamma - 1$  to provide a conservative range of values at high redshift.

The spatial correlation function slope,  $\gamma$ , is the same as that used in the power-law fit to the angular correlation function (where the magnitude of the slope is  $\gamma - 1$ ), so we measure this parameter

through the  $w(\theta)$  function. The amplitude  $A$  of  $w(\theta)$  has been expressed as a function of  $r_0$  (in comoving coordinates) in the literature (Overzier et al. 2003; Kovač et al. 2007; Kim et al. 2011; Elyiv et al. 2012) as follows:

$$A = r_0^\gamma H_\gamma \left( \frac{H_0}{c} \right) \frac{\int_0^\infty N^2(z)(1+z)^{\gamma-(3+\epsilon)} \chi^{1-\gamma}(z) E(z) dz}{\left[ \int_0^\infty N(z) dz \right]^2}, \quad (9)$$

where  $H_\gamma$  is related to the Gamma function,  $H_\gamma = \Gamma(\frac{1}{2})\Gamma(\frac{\gamma-1}{2})/\Gamma(\frac{\gamma}{2})$ ,  $N(z)$  is the redshift distribution and  $\chi(z)$  is the comoving line-of-sight distance to an object at a redshift  $z$ :

$$\chi(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}. \quad (10)$$

Here,  $H_0$  is the Hubble constant and  $E(z)$  is the function used to describe the cosmological expansion history:

$$E(z) = [\Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0}]^{\frac{1}{2}}. \quad (11)$$

Equation (9) may simply be inverted to give the comoving correlation length,  $r_0$  as a function of (i) redshift distribution, (ii) correlation function slope and (iii) angular clustering amplitude.

We employ redshift distributions as measured for the GAMA matched sources, and for the simulated SKADS sources, as detailed in Section 2.5. We also emphasize that using the SKADS  $N(z)$  allows us to account for sensitivity variations across the FIRST survey as we able to remove the correct number of sources of a given flux and redshift (as described in Section 2.6).

#### 4.1 Results of Limber inversion

We have measured  $w(\theta)$  for the six samples described above, and fitted them with a power law (see Table 4). Given the dependence on  $\epsilon$ , the clustering index of choice,  $r_0$  is shown for the comoving ( $\epsilon = \gamma - 3$ ) and linear ( $\epsilon = \gamma - 1$ ) clustering models. The latter consistently gives lower values for  $r_0$ , with the unshown stable ( $\epsilon = 0$ ) model values being between the two, but these differences tend to be comparable with the associated errors. The results of the comoving clustering model will be quoted hereafter, unless stated otherwise.

The large FIRST sample yields a  $w(\theta)$  consistent with the literature, where we find a slope of  $\gamma - 1 = 0.82 \pm 0.02$ , the collapsing of close pairs successfully having removed evidence of a steeper, small-scale power-law component. By assuming the same redshift distribution as a similar SKADS catalogue (adjusted in accordance with the noise variations in the FIRST survey, see Section 2.6), we find a correlation length of  $8.20^{+0.41}_{-0.42} h^{-1}$  Mpc at a median redshift of  $z = 1.21$ .

We are able to infer the clustering properties from our real data within the GAMA fields at three different epochs utilizing the three subsamples described in Section 2.5 and Fig. 4. These three subsamples have distributions of median  $z = 0.48, 1.21$  and  $1.55$  for the matched, complete and unmatched radio samples, respectively.

In each case, the  $w(\theta)$  power law is found to be significantly steeper ( $\gamma > 2$ ) than for the larger FIRST sample, albeit less well constrained owing to the smaller sample sizes. This steeper power law is borne out in the analysis where we find that the ‘FIRST–GAMA fields’ sample has a larger clustering length than the larger FIRST sample with the same  $N(z)$ . The trend in these three subsamples, however, is for increasing  $r_0$  with redshift. This is confirmed by analysing subsamples of the cross-matched objects. Small number counts limit us to taking a simple high- and low-redshift approach,

but even within this redshift range of  $z \sim 0.3$  to  $\sim 0.65$ ,  $r_0$  is seen to increase.

In order to assess the effect of cosmic variance on our results over the GAMA fields, we also find  $w(\theta)$  and  $r_0$  for 20 separate  $210 \text{ deg}^2$  patches of the FIRST survey over a large range of declinations ( $-5^\circ < \delta < +55^\circ$ ), assuming the SKADS  $N(z)$  for each. We find a mean and standard deviation of correlation lengths from these subsamples to give  $r_0 = 9.41 \pm 1.57 h^{-1}$  Mpc, lying between the two values found for our GAMA sample and for the larger FIRST sample at  $z = 1.21$ , and placing both within this cosmic variance error.

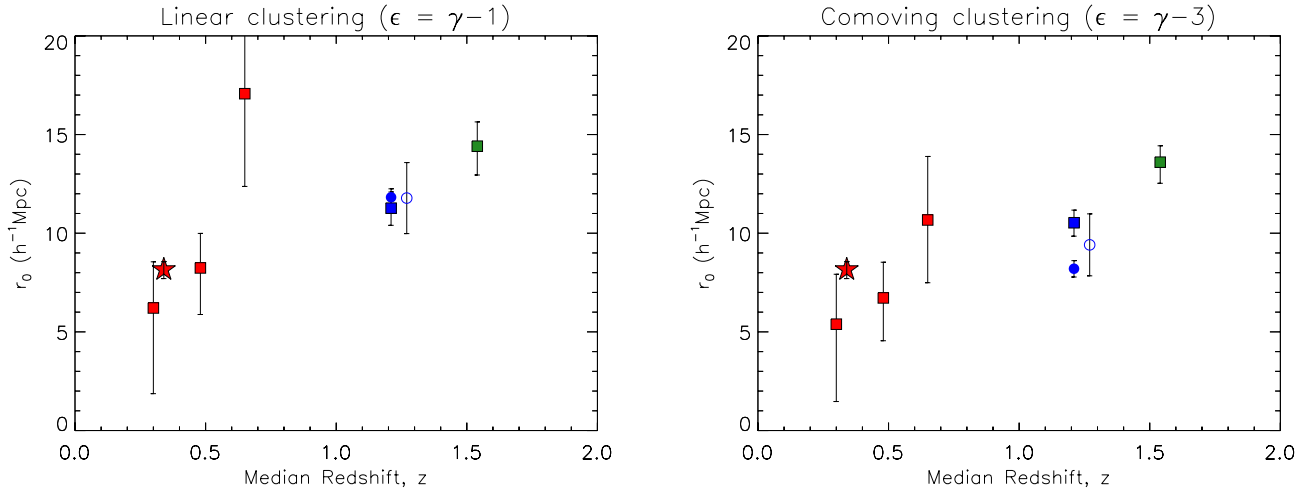
#### 4.2 Discussion

Previous results using the angular clustering approach, without individual redshifts, Blake & Wall (2002a) find a clustering length for  $S > 10 \text{ mJy}$  NVSS sources of  $r_0 \sim 6 h^{-1}$  Mpc (independent of flux density limit) under the assumptions of  $N(z)$  calculated from luminosity functions of Dunlop & Peacock (1990) and an Einstein–de Sitter (EdS) cosmological model. Similarly, Overzier et al. (2003) use NVSS and similarly derived  $N(z)$  from Dunlop & Peacock (1990) finding  $r_0 \sim 5 \pm 1 h^{-1}$  Mpc for flux density between 3 and 40 mJy. Both papers probe clustering at  $z \sim 1$  and, while they differ slightly from one another, they are both considerably lower than our findings even with our lower flux limit diluting the clustering due to AGN. It is important to note, however, that the Dunlop & Peacock models used are poorly constrained at these low flux density limits due to a lack of volume coverage at low redshift in their combined surveys. Improvements in deep surveys have allowed Wilman et al. (2008) to better constrain low-power AGN and star-forming galaxies in the SKADS redshift distribution for use at the mJy level.

Any underestimation of  $r_0$  could partly be a result of having fixed the angular clustering power law to a slope of  $-0.8$  as is widely observed for normal galaxies, although even our larger FIRST sample with  $\gamma - 1 = 0.77$  appears to exhibit stronger clustering. This may be simply to reduce the power-law fitting to a one-parameter problem, but a steeper slope ( $\sim 1.0$ – $1.4$ ) may give a higher correlation length than found under the assumption of a universal slope of  $-0.8$ . It is likely, however, that these differing results are due mostly to our lower flux density limits and slightly higher median redshift.

Indeed, while our  $w(\theta)$  fit for the larger FIRST sample does yield the assumed  $(\gamma - 1) = 0.8$ , Cress et al. (1996) find a higher value, similar to our smaller subsamples, of  $(\gamma - 1) \sim 1.1$  yielding a larger correlation length at the same  $z \sim 1$  of  $r_0 \sim 10 h^{-1}$  Mpc, more in keeping with our results. Furthermore, this epoch and corresponding correlation length is in close agreement with observations of  $z \sim 1$  EROs using the same  $\epsilon = \gamma - 3$  clustering model (e.g. Daddi et al. 2001, 2002; McCarthy et al. 2001; Roche et al. 2002). This ties in with the suggestion by Willott et al. (2001) that EROs and high-redshift galaxies are different evolutionary stages of the same galaxy population, which in turn may be the progenitors of local bright ellipticals.

Two studies of spectroscopic samples of radio sources associated with luminous red galaxies (LRGs) by Wake et al. (2008) and Fine et al. (2011) investigate samples at a similar range of redshifts below  $z = 0.8$  by different methods. Wake et al. (2008) fit a halo occupation distribution (HOD) model to the projected spatial correlation function of radio-detected ( $L_{1.4} \gtrsim 10^{24.2} \text{ W Hz}^{-1}$ ) LRGs at  $z \sim 0.55$  to give  $r_0 \sim 12 h^{-1}$  Mpc, slightly greater than we find (Fig. 7) due to their selection of the most optically luminous, more powerful AGN sources. Fine et al. (2011) show the evolution in this



**Figure 7.** Correlation length for the observed radio samples described in Table 4 as calculated for two different clustering indices. Filled square symbols correspond to samples over the GAMA/SDSS/UKIDSS fields, while the circle symbols refer to the wider FIRST survey and the star symbols correspond to the spectroscopic sample of FIRST matches in GAMA. Red, green and blue colours refer to matched, unmatched and total samples, respectively (cf. Fig. 4). The open circle beside the points at  $z \sim 1.2$  shows the mean value and error associated with 20 independent patches of the FIRST survey, providing an estimate of the sample variance within a field size of  $210 \text{ deg}^2$ .

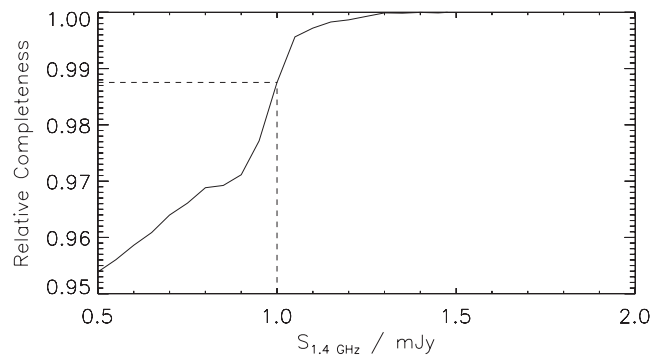
clustering by adding LRG samples from SDSS ( $z \sim 0.35$ ) and AAOmega ( $z \sim 0.68$ ). Over this very similar redshift range to our optically identified radio sources, these authors find  $r_0$  using angular auto- and cross-correlation functions which are consistent with no redshift evolution, falling in the range of  $\sim 10\text{--}12 \text{ h}^{-1} \text{ Mpc}$ . While we find a similar correlation length towards higher redshift, this result is contrary to the trend we observe with the GAMA/SDSS radio sources. Again, this could simply be a result of the different optical selection favouring considerably more powerful AGN at lower redshift, whereas our magnitude limit allows more of the low optical luminosity sources into our sample.

Looking at the AGN population from the other end of the electromagnetic spectrum, Elyiv et al. (2012) use a  $N(z)$  derived from a range of luminosity functions for AGN detected in X-rays with the *XMM*-Large Scale Structure (LSS), recovering the ubiquitous slope of  $-0.8$  for the soft band sources, but  $-1.0$  for the hard band. Their hard band sources exhibit a clustering length of  $r_0 \simeq 10 \pm 1 \text{ h}^{-1} \text{ Mpc}$  at  $z \sim 1$  using a clustering index of  $\epsilon = -1.2 \simeq \gamma - 3$  roughly corresponding to comoving clustering evolution. Accounting for the slightly lower redshift, this agrees well with a FIRST radio population expectedly dominated by AGN emission.

#### 4.2.1 Surface density variations and sample variance

Blake & Wall (2002b) have shown some declination dependence in the surface density of FIRST sources, with a dearth at lower declination of fainter radio sources in particular. We find this to be an 8–10 per cent underdensity for sources fainter than 10 mJy with  $|\delta| < 3^\circ$ , a range encompassing the GAMA fields. Given that the source densities remain consistent with expectations from SKADS, this density variation may be a result of increased noise in this region of the survey, sample variance or interferometric effects arising from observing at low declination. Accounting for any variation in the noise values across the FIRST survey, however, finds only a small difference ( $\sim 1$  per cent at the flux density limit) in completeness of the equatorial fields and the wider survey (see Fig. 8).

As a result, we might expect the clustering statistics to reflect a larger proportion of radio-loud AGN rather than low-redshift star-



**Figure 8.** Estimated ratio of completeness curves for the FIRST survey over the three equatorial GAMA/SDSS fields and a large area of the northern sky. Higher rms noise values at lower declinations make the GAMA fields less complete at the lowest flux densities.

forming galaxies which become more dominant at the mJy level. We demonstrate this effect by calculating  $r_0$  and bias for 20 patches of FIRST, each  $210 \text{ deg}^2$  in area, over a declination range between  $-6^\circ$  and  $+54^\circ$ . The open circles in Fig. 7, slightly offset in  $z$  for clarity, represent the mean and standard error on these quantities. The GAMA fields yield results at the high end of what bias we might find based on the surface density fluctuation, while the larger FIRST sample at  $\delta > 0$  is towards the lower end, however, we note that both are consistent with the expected variance with patches of  $210 \text{ deg}^2$ .

## 5 SPATIAL CORRELATION FUNCTION

In addition to calculating the spatial correlation length to high redshift using the deprojected angular correlation function, we are also able to determine the spatial correlation function directly for those sources with spectroscopic redshifts. Unlike the angular case, calculating spatial separations between galaxies requires information about the underlying cosmology and is dependent on the dynamics of the clustered systems. On face value, the only extra data required

for our catalogue is a line-of-sight distance, which we calculate based on observed redshifts. Given the comoving distances to two objects,  $\chi_1$  and  $\chi_2$  (from equation 10), and their angular separation,  $\theta$ , their comoving spatial separation is given by

$$r = (\chi_1^2 + \chi_2^2 - 2\chi_1\chi_2 \cos \theta)^{\frac{1}{2}}, \quad (12)$$

assuming a flat cosmology (see Liske 2000 for more general expressions).

The spatial two-point correlation function is usually fitted by a single power law over a significant range of separations:

$$\xi(r, z) = \left( \frac{r_0(z)}{r} \right)^\gamma, \quad (13)$$

where  $r_0(z) = r_0(1+z)^{-\frac{3+\epsilon}{\gamma}}$ , incorporating the clustering evolution described in Section 4.

### 5.1 Projected correlation function

By calculating galaxy distances from their redshifts and a cosmological model as in equations (10) and (12), we actually find the *redshift-space* correlation function,  $\xi(s)$ , which is systematically different from the true *real-space* correlation function  $\xi(r)$ . The difference occurs because measured redshifts are not entirely the result of cosmological expansion but they are also affected by radial peculiar velocity components of the source distorting the observer's view (the Kaiser effect, Kaiser 1987, and 'Fingers of God'). These effects make it difficult to find the real-space clustering parameters  $r_0$  and  $\gamma$  from direct  $\xi(s)$  measurements. These can be found directly through the projected correlation function  $\Xi(\sigma)$  or indirectly through deprojecting the angular correlation function  $w(\theta)$  (Section 4).

Redshift-space distortions affect our measurements of galaxy separations only in the direction of their lines of sight. This can be ameliorated by calculating the correlation function as a function of the line-of-sight separation,  $\pi$ , and the transverse separation,  $\sigma$ . Integrating over the  $\pi$  coordinate gives the projected correlation function, which is a function of the ostensibly real-space  $\sigma$  coordinate:

$$\Xi(\sigma) = 2 \int_0^\infty \xi(\sigma, \pi) d\pi. \quad (14)$$

For practical purposes we must impose our own upper limit on  $\pi$  for the integration in order to strike a compromise between capturing the full clustering signal and avoiding the introduction of excessive noise at the highest separations. Hawkins et al. (2003), for example, find their results insensitive to  $\pi_{\max} > 60 h^{-1}$  Mpc. We use a limit of  $\pi_{\max} = 50 h^{-1}$  Mpc in this work, avoiding considerable noise in  $\xi(\pi > 50, \sigma)$ . This function, free of redshift-space effects can be related to the real-space correlation function (Davis & Peebles 1983):

$$\frac{\Xi(\sigma)}{\sigma} = \frac{2}{\sigma} \int_0^\infty \frac{r \xi(r)}{\sqrt{r^2 - \sigma^2}} dr. \quad (15)$$

Assuming a power law form for  $\xi(r)$  as in equation (13), we can fit  $\Xi(\sigma)/\sigma$  to find the real-space correlation length,  $r_0$ , and power-law slope,  $\gamma$ :

$$\frac{\Xi(\sigma)}{\sigma} = \left( \frac{r_0}{\sigma} \right)^\gamma H_\gamma. \quad (16)$$

Naturally, the requirement of having spectroscopic redshifts to accurately determine the galaxy pair separations limits the use of

**Table 5.** Clustering parameters found from the redshift-space correlation function  $\xi(s)$  and the projected correlation function  $\Xi(\sigma)$  of the 1635 source sample of radio galaxies with spectroscopic redshifts from GAMA.

Function	Range ( $h^{-1}$ Mpc)	$s_0/r_0$ ( $h^{-1}$ Mpc)	$\gamma$	$b(z = 0.34)$
$\xi(s)$	$2 < s < 300$	$7.80^{+0.40}_{-0.53}$	$1.34^{+0.05}_{-0.05}$	$1.72^{+0.08}_{-0.08}$
$\Xi(\sigma)$	$\sigma < 50$	$8.16^{+0.40}_{-0.46}$	$1.60^{+0.04}_{-0.05}$	$1.92^{+0.10}_{-0.11}$

this method to those 1635 radio sources with GAMA spectroscopy. Furthermore, by binning galaxy pairs in two dimensions and then integrating, we sacrifice signal-to-noise ratio in any given  $\sigma$  bin, limiting the scope of the available data. We most likely underestimate the errors associated with  $\Xi(\sigma)$ , however, as the errors on the numerous  $\pi$  bins being integrated are treated as independent.

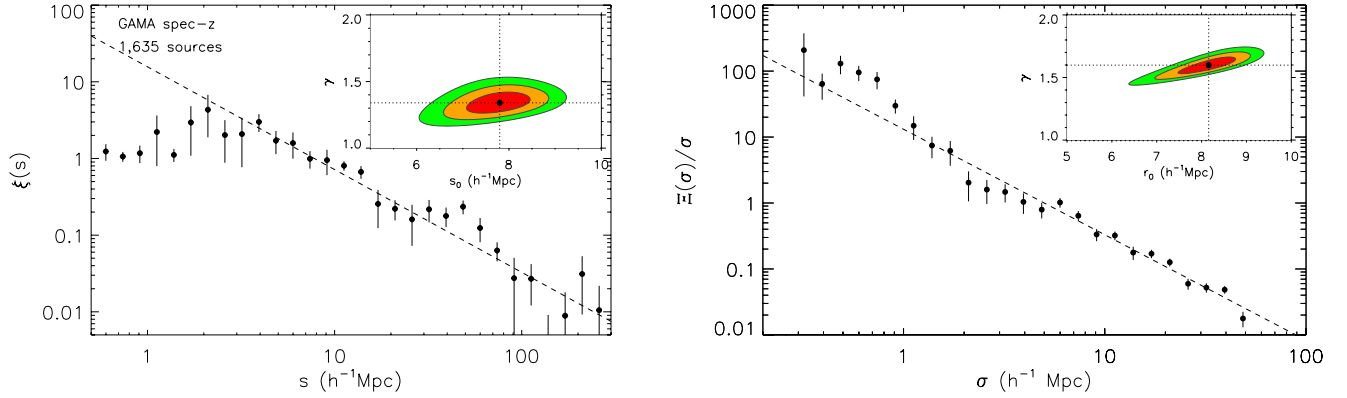
Given the advantages gained by having spectroscopic redshifts, we have measured the redshift-space and projected correlation functions of those radio sources with optical counterparts in the GAMA survey itself (Table 5 and Fig. 9). For this sample at  $z \sim 0.34$ , we find a redshift-space correlation length  $s_0 = 7.80^{+0.40}_{-0.53} h^{-1}$  Mpc with  $\gamma \sim 1.34$ , and correcting for redshift-space distortions, for the projected correlation function we find  $r_0 = 8.16^{+0.40}_{-0.46} h^{-1}$  Mpc with  $\gamma \sim 1.60$ .

### 5.2 Results and discussion

Our more direct spatial clustering measures at low redshift using the spectroscopic sample appear to agree with the results from the deprojection of the angular correlation function,  $w(\theta)$ , albeit with the latter constraints on  $r_0$  and bias being much weaker. The results of this spectroscopic sample lend themselves to comparison with the work of Brand et al. (2005) and Magliocchetti et al. (2004). Brand et al. (2005) directly measure the spatial correlation function (although ignoring redshift-space distortions) of 268 radio galaxies in  $165 \text{ deg}^2$  of the Texas-Oxford NVSS Structure (TONS) survey with flux density  $S_{1.4} > 3 \text{ mJy}$  and an optical limit of  $R \lesssim 19.5$ . The slightly higher radio flux limit removes some of the sources associated with star formation, but the brighter optical limits favour low-luminosity radio galaxies at lower redshift, with the median of the sample at  $z \sim 0.3$ . They find  $r_0 = 6.1 \pm 1.1 h^{-1}$  Mpc and assume linear ( $\epsilon = \gamma - 1$ ) clustering model in good agreement with our corresponding measure with the 'FIRST-GAMA-matched- $z < 0.5$ ' sample.

Magliocchetti et al. (2004) make joint use of FIRST and 2dFGRS to a radio flux density limit of 1 mJy and an optical limit of  $b_1 < 19.37$ , working with a sample of 820 radio sources with redshifts  $0.01 < z < 0.3$  over a larger area of  $\sim 375 \text{ deg}^2$ . Their value of  $r_0 \sim 4.7 \pm 0.7 h^{-1}$  Mpc is lower than for our sample, but perhaps by no more than expected given their lower redshifts and preferential selection of optically brighter sources. This effect is mitigated by their subsequent selection of only those sources whose spectra have signatures of AGN activity. This increases the redshift from  $z \sim 0.10$  to 0.13, and the correlation length to  $r_0 \sim 7.6 \pm 0.8 h^{-1}$  Mpc, similar to that of our sample  $z \sim 0.34$ , shown in Fig. 7. We would expect to observe clustering between these two values, albeit boosted due to our sample having a higher median redshift. The close agreement with these past surveys also suggests that our analysis accounting for the noise variations in the GAMA fields is robust.





**Figure 9.** The redshift-space correlation function (left) and projected spatial correlation function (right) for the GAMA spectroscopically identified objects with bootstrap resampling errors. The dashed line shows the best-fitting power law (over the ranges  $2 < s < 300 h^{-1} \text{ Mpc}$  and  $\sigma < 50 h^{-1} \text{ Mpc}$ , respectively) and the inset plot shows  $\chi^2$  parameter fits at 68, 90 and 95 per cent confidence levels.

## 6 MASS BIAS

The differences in clustering of different classes of extragalactic objects and the background matter distribution motivate the use of some bias parameter,  $b$ , as introduced by Kaiser (1984) and Bardeen et al. (1986):

$$b^2(z) = \frac{\xi_{\text{gal}}(r, z)}{\xi_{\text{DM}}(r, z)}, \quad (17)$$

where the numerator and denominators are the galaxy and dark matter correlation functions, respectively.

The bias parameter can be determined directly if both the real- and redshift-space forms of the correlation function are known. The ratio of the two can be approximated by a quadratic in the infall parameter  $\beta \propto b(z)$  (Kaiser 1987). Therefore, with a good estimate of the real- and redshift-space correlation functions, one may calculate the ratio of the two and thereby the bias.

We are able to infer the real-space correlation length  $r_0$  from the angular correlation function (Section 4) or the projected correlation function (Section 5.1). The bias parameter (as a function of redshift) may be defined as equation (17) with  $r = 8 h^{-1} \text{ Mpc}$ . As described in equation (13) the numerator can be written as

$$\xi_{\text{gal}}(8, z) = \left[ \frac{r_0(z)}{8} \right]^\gamma. \quad (18)$$

The corresponding function for the denominator is given by Peebles (1980) as

$$\xi_{\text{DM}}(8, z) = \sigma_8^2(z) / J_2, \quad (19)$$

where  $J_2 = 72 / [(3 - \gamma)(4 - \gamma)(6 - \gamma)2^\gamma]$  and the parameter  $\sigma_8^2$  is the dark matter density variance in a comoving sphere of radius  $8 h^{-1} \text{ Mpc}$ . The combination of these equations gives the scale-independent evolution of bias with redshift, given only the correlation length and slope:

$$b(z) = \left[ \frac{r_0(z)}{8} \right]^{\gamma/2} \frac{J_2^{1/2}}{\sigma_8 D(z)/D(0)}. \quad (20)$$

In our subsequent analysis, the redshift value is assumed to be the median of the distribution of objects.

### 6.1 Results

The bias inferred from the angular correlation function is shown in Table 4 for both  $\epsilon$  values corresponding to linear and comoving

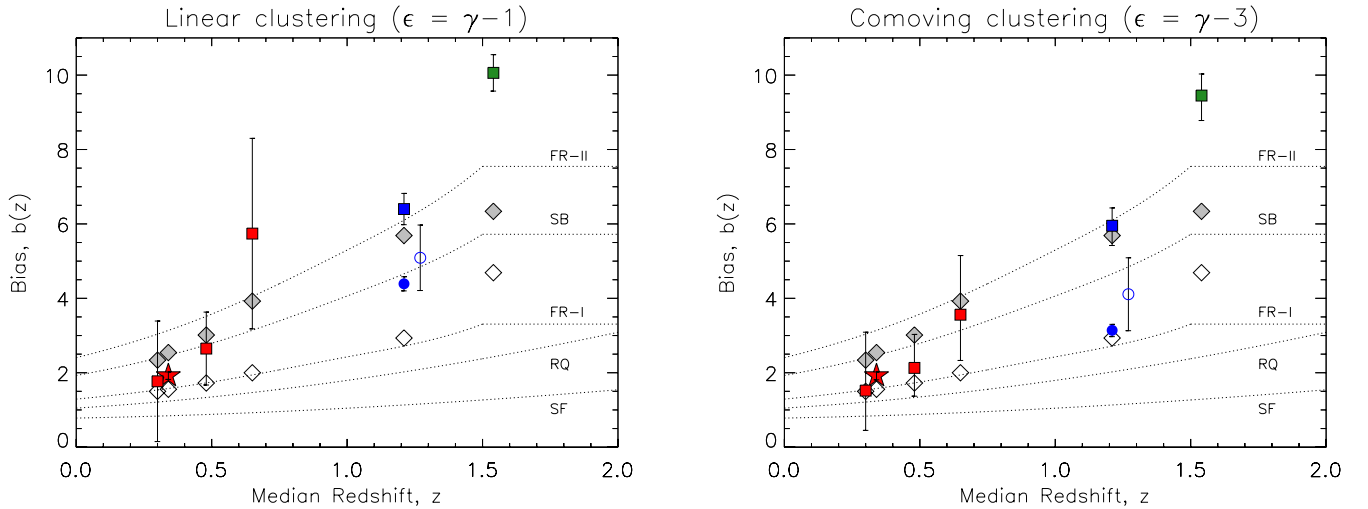
clustering evolution. These results are compared in Fig. 10 with the bias model by Mo & White (1996) used by Wilman et al. (2008) in populating dark matter haloes in the SKADS simulations. They assign a particular halo mass to each source type in the simulation: FRI, FRII and radio-quiet AGN, normal star-forming galaxies and starbursts, with assumed halo masses of  $10^{13}$ ,  $10^{14}$ ,  $3 \times 10^{12}$ ,  $10^{11}$  and  $5 \times 10^{13} h^{-1} \text{ M}_\odot$ , respectively, and impose a plateau for each model at high redshifts above which the assumption of a fixed halo mass breaks down. The model  $b(z)$  for each of these source types is denoted by a black dotted line, showing the stronger bias for those objects residing in more massive haloes.

By weighting the SKADS sources by the ratio between the sample and SKADS redshift distributions, we have estimated the relative proportions of the population masses. Calculating the model bias of the sample as the weighted mean of the biases of each individual population, we show a predicted bias at the five different redshifts probed in Fig. 10 (open diamonds), accounting for the change in galaxy halo mass being observed. These follow the same trend as our measurements, with the bias rising more steeply with redshift than any individual model population owing to the increasing AGN fraction towards higher  $z$ . Our measurements over the GAMA fields agree qualitatively with the model, but with those associated only with the GAMA fields exceeding the predictions to varying degrees. For those sources with measured redshifts, the difference is within the stated errors, but at higher redshifts our measurements exceed the SKADS bias values by  $\sim 3\sigma$  (we return to this point in Section 6.2).

The spatial correlation function of spectroscopic sources provides one additional data point for the bias at a redshift of  $z = 0.34$ , similar to the  $z < 0.5$  subsample of combined spectroscopic and photometric matches. Table 5 shows  $r_0$  and  $s_0$  derived from  $\xi(s)$  and  $\Xi(\sigma)$ , respectively, and the inferred bias (where redshift-space effects are wrongly ignored in the case of the former). We find a bias of  $\sim 1.7$  for the redshift-space calculation of the bias, with a real-space equivalent of  $\sim 1.9$  at  $z \sim 0.34$ . The results from  $\Xi(\sigma)$  are shown in Fig. 10 for comparison with the angular clustering results.

### 6.2 Discussion

Within our matched sample and incorporating simulated data, we observe the increasing of  $b(z)$  in line with models corresponding to fixed masses (Mo & White 1996; Matarrese et al. 1997), although we



**Figure 10.** Linear bias parameter for the observed radio samples described in Table 4 as calculated for two different clustering indices, with symbols as defined in Fig. 7. Overplotted are lines showing the model bias evolution based on the halo masses assigned in the SKADS simulation of individual source populations: FRI, FRII and radio-quiet (RQ) AGN, starbursts (SB) and normal star-forming (SF) galaxies. Open diamonds corresponding to each point mark the aggregate bias expected of the samples, assuming relative population abundances and masses used in redshift-matched SKADS samples. The grey filled diamonds show this same model bias if we increase the assumed halo mass of FRI sources to equal that of the FRII sources.

find a relatively high value for the high-redshift sources, particularly when using the linear clustering parameter. Whilst the GAMA data points do not place especially tight constraints on this evolution, and the well-constrained point from  $\Xi(\sigma)$  is limited to one epoch, the assumption of the redshift distributions for the full and unmatched subsamples probe redshifts up to  $z = 1.55$  and give more shape to the evolution. The agreement is good at lower redshifts, the GAMA-matched results (red) lying  $< 1\sigma$  from the model prediction as well as agreeing with the 2SLAQ radio-detected LRGs examined by Wake et al. (2008) ( $b \sim 3.0$ ). The higher redshift points, however, significantly exceed the values prescribed in the SKADS simulation. The wider FIRST sample is an exception, falling significantly below the trend displayed by the narrower samples, but closely matching the Mo & White bias prescription from SKADS. On face value, this apparent excess clustering suggests a greater proportion of AGN observed than was assumed in the simulations, or perhaps a considerable underestimate of the typical halo masses of any or all of the galaxy types considered.

Our unmatched FIRST sources, especially, in the GAMA field are more strongly clustered than would be expected by simply subtracting the cross-matched  $N(z)$  from the SKADS  $N(z)$  and assuming the same population as a matching SKADS sample. This means either that the assumed redshift distribution is skewed towards low redshift, or the fraction of more massive galaxies are higher than expected. Inspection of Fig. 4 shows that our assumed  $N(z)$  for this point contains a significant fraction of  $z < 0.5$  sources where one would expect to observe a greater proportion of less massive, star-forming galaxies rather than the more massive AGN. Given that if these low-redshift objects were detected in FIRST, we would expect them to have optical counterparts in SDSS/GAMA (cf. almost complete cross-identification at  $z \sim 0.6$  in Fig. 4), it is perhaps unrealistic to assume these missing low- $z$  sources remain in our unmatched sample. The cause of this large fraction of unmatched sources is discussed in Section 2.5, however, to determine the effect that these sources have on our measured clustering length and bias we implement cuts on the SKADS redshift distribution, where we remove all sources at  $z < 0.05$  and  $< 0.1$  from the simulations. This is a coarse way of simulating the combined effect of the resolution

bias and the removal of very low redshift sources with photometric redshifts with  $z < 0.002$ . We then recalculate the correlation length and bias using these new redshift distributions and find that the correlation length and bias increase significantly ( $\sim 25$ – $50$  per cent increase in bias). Although subject to many uncertainties, this may point to a higher proportion of highly biased objects in the radio source population at high redshift.

One of the main uncertainties in the SKADS simulation is indeed the evolution and bias of the FRI-type objects which dominate the source counts at the flux density limit of the FIRST survey at  $z > 0.5$ . In the SKADS simulation the FRIs are less biased than the FRIIs, however, it is becoming clear that there is a large overlap in the stellar mass distributions between the generally less radio luminous FRIs and their FRII counterparts at high redshifts (e.g. McLure et al. 2004; Herbert et al. 2011), thus it is possible that the bias prescription for the FRI sources in the SKADS simulation is underestimated. If we assign a similar halo mass to the FRI as for the FRIIs, then we find that the expected average bias of radio sources in our sample ( $z \sim 1.21$ ) to be in the region of  $b = 5.7$ , this is much more closely aligned with our measured value of  $b = 5.95^{+0.48}_{-0.53}$ .

Deeper surveys over smaller areas will be able to address this issue better than the relatively shallow FIRST data. Indeed, deep multiwavelength surveys are now beginning to tackle the question of redshift evolution (e.g. Smolčić et al. 2009; McAlpine & Jarvis 2011; Simpson et al. 2012; McAlpine, Jarvis & Bonfield 2013) and with slightly more area could measure the clustering length.

## 7 CONCLUSIONS

Using radio observations from the FIRST survey and optical/infrared data (SDSS/UKIDSS LAS) over the GAMA survey field area, we have obtained a cross-matched sample of radio galaxies with optical host galaxy redshifts ( $\sim 42$  per cent spectroscopic). We have measured the redshift-space and projected correlation functions of these spectroscopic identifications from GAMA, as well as the angular correlation function of radio sources over the extended

GAMA survey area. Assuming parent redshift distributions from SKADS (Wilman et al. 2008), we have inferred the spatial correlation length  $r_0$  and the mass bias  $b(z)$  for the matched radio sources and the radio sources without optical identifications, extending our redshift range up to  $z \sim 1.6$ . The results can be summarized as follows.

(i) The projected correlation function of FIRST sources with GAMA spectroscopic counterparts provides a well-constrained correlation length and linear bias of  $r_0 = 8.16^{+0.40}_{-0.46} h^{-1}$  Mpc and  $b = 1.92^{+0.10}_{-0.11}$  at  $z \sim 0.34$ .

(ii) The angular correlation function follows a power law but we find a steeper slope of  $\sim -1.2$  compared with that found for other classes of galaxy and often assumed in the literature.

(iii) Our cross-matched sample yields a spatial correlation length of  $r_0 \sim 8.5$  to  $\sim 10.7 h^{-1}$  Mpc at  $0.3 < z < 0.65$  (assuming stable clustering evolution). Adding the assumption of the SKADS parent  $N(z)$ , we measure the clustering length up to median redshift  $z \sim 1.55$ , where we find  $r_0 \sim 12.2 h^{-1}$  Mpc.

(iv) We measure the bias as a function of redshift across the subsamples finding it to increase from  $b(z = 0.30) \sim 2.8$  to  $b(z = 1.55) \sim 9.2$ . These values were compared with predicted values from a model assuming population fractions and masses from SKADS and were in qualitative agreement but exceeding the prescribed values at high redshift. This is most probably due to a combination of surface density fluctuation in the FIRST survey, leading to a shortfall of sources at  $\sim 1$  mJy which in turn biases our sample towards a higher fraction of strongly clustered AGN at high redshifts and/or potential inaccuracies in the halo masses of particular radio subpopulations used in SKADS.

(v) If we assign a similar halo mass to the FRI sources as assumed for the FRII sources in our radio survey then we find that we can reproduce our bias value at high redshift. A more highly biased FRI population may in turn lead to a highly biased tracer of the high-redshift Universe for cosmological applications with radio surveys (e.g. Camera et al. 2012; Raccanelli et al. 2012).

While we place modest constraints on the clustering evolution of mJy radio sources and suffer from limited cross-identification with the optical surveys, our use of a well-constrained redshift distribution from SKADS allows us to extend our redshift range to  $z > 1$ . This highlights some potentially important discrepancies with a simple model bias, implying stronger clustering at higher redshifts than is expected from the fixed halo masses assumed in the SKADS simulations through some combination of a greater proportion of massive AGN than previously thought and/or an increased typical halo mass being observed. In a future paper, we analyse deep field data to help confirm our high-redshift bias estimates.

## ACKNOWLEDGEMENTS

GAMA is a joint European–Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including GALEX MIS, VST KIDS, VISTA VIKING, WISE, *Herschel*-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

## REFERENCES

- Afonso J., Mobasher B., Koekemoer A., Norris R. P., Cram L., 2006, *AJ*, 131, 1216
- Bahcall N. A., Soneira R. M., 1983, *ApJ*, 270, 20
- Baldry I. K. et al., 2010, *MNRAS*, 404, 86
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
- Becker R. H., White R. L., Helfand D. J., 1995, *ApJ*, 450, 559
- Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, *MNRAS*, 362, 25
- Blake C., Wall J., 2002a, *MNRAS*, 329, L37
- Blake C., Wall J., 2002b, *MNRAS*, 337, 993
- Blake C. et al., 2011, *MNRAS*, 418, 1707
- Bondi M. et al., 2003, *A&A*, 403, 857
- Brand K., Rawlings S., Hill G. J., Tufts J. R., 2005, *MNRAS*, 357, 1231
- Camera S., Santos M. G., Bacon D. J., Jarvis M. J., McAlpine K., Norris R. P., Raccanelli A., Röttgering H., 2012, *MNRAS*, 427, 2079
- Ciliegi P. et al., 2005, *A&A*, 441, 879
- Cole S. et al., 2005, *MNRAS*, 362, 505
- Colless M. et al., 2001, *MNRAS*, 328, 1039
- Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, *AJ*, 115, 1693
- Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L., 1996, *ApJ*, 473, 7
- Croom S. M., Shanks T., Boyle B. J., Smith R. J., Miller L., Loaring N. S., Hoyle F., 2001, *MNRAS*, 325, 483
- Croom S. M. et al., 2005, *MNRAS*, 356, 415
- Daddi E., Broadhurst T., Zamorani G., Cimatti A., Röttgering H., Renzini A., 2001, *A&A*, 376, 825
- Daddi E. et al., 2002, *A&A*, 384, L1
- Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
- de Zotti G., Persic M., Franceschini A., Danese L., Palumbo G. G. C., Boldt E. A., Marshall F. E., 1990, *ApJ*, 351, 22
- Driver S. P. et al., 2011, *MNRAS*, 413, 971
- Dunlop J. S., Peacock J. A., 1990, *MNRAS*, 247, 19
- Eales S. et al., 2010, *PASP*, 122, 499
- Eisenstein D. J. et al., 2005, *ApJ*, 633, 560
- El Boucheffry K., Cress C. M., 2007, *Astron. Nachr.*, 328, 577
- Elyiv A. et al., 2012, *A&A*, 537, A131
- Fine S., Shanks T., Nikoloudakis N., Sawangwit U., 2011, *MNRAS*, 418, 2251
- Foucaud S. et al., 2007, *MNRAS*, 376, L20
- Giannantonio T., Scranton R., Crittenden R. G., Nichol R. C., Boughn S. P., Myers A. D., Richards G. T., 2008, *Phys. Rev. D*, 77, 123520
- Gonzalez-Solares E. A. et al., 2005, *MNRAS*, 358, 333
- Grazian A. et al., 2006, *A&A*, 449, 951
- Groth E. J., Peebles P. J. E., 1977, *ApJ*, 217, 385
- Guzzo L., Strauss M. A., Fisher K. B., Giovanelli R., Haynes M. P., 1997, *ApJ*, 489, 37
- Hawkins E. et al., 2003, *MNRAS*, 346, 78
- Herbert P. D., Jarvis M. J., Willott C. J., McLure R. J., Mitchell E., Rawlings S., Hill G. J., Dunlop J. S., 2011, *MNRAS*, 410, 1360
- Hill D. T. et al., 2011, *MNRAS*, 412, 765
- Hopkins P. F., Cox T. J., Hernquist L., Narayanan D., Hayward C. C., Murray N., 2013, *MNRAS*, 430, 1901
- Ivashchenko G., Zhdanov V. I., Tugay A. V., 2010, *MNRAS*, 409, 1691
- Jarvis M. J., Rawlings S., Eales S., Blundell K. M., Bunker A. J., Croft S., McLure R. J., Willott C. J., 2001, *MNRAS*, 326, 1585
- Jarvis M. J. et al., 2010, *MNRAS*, 409, 92
- Kaiser N., 1984, *ApJ*, 284, L9
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kim J.-W., Edge A. C., Wake D. A., Stott J. P., 2011, *MNRAS*, 410, 241
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Kooiman B. L., Burns J. O., Klypin A. A., 1995, *ApJ*, 448, 500
- Kovač K., Somerville R. S., Rhoads J. E., Malhotra S., Wang J., 2007, *ApJ*, 668, 15
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64

- Lawrence A. et al., 2007, *MNRAS*, 379, 1599
- Limber D. N., 1953, *ApJ*, 117, 134
- Ling E. N., Barrow J. D., Frenk C. S., 1986, *MNRAS*, 223, 21P
- Liske J., 2000, *MNRAS*, 319, 557
- Loan A. J., Wall J. V., Lahav O., 1997, *MNRAS*, 286, 994
- McAlpine K., Jarvis M. J., 2011, *MNRAS*, 413, 1054
- McAlpine K., Smith D. J. B., Jarvis M. J., Bonfield D. G., Fleuren S., 2012, *MNRAS*, 423, 132
- McAlpine K., Jarvis M. J., Bonfield D. G., 2013, *MNRAS*, 436, 1084
- McCarthy P. J. et al., 2001, *ApJ*, 560, L131
- McEwen J. D., Vielva P., Hobson M. P., Martínez-González E., Lasenby A. N., 2007, *MNRAS*, 376, 1211
- McLure R. J., Willott C. J., Jarvis M. J., Rawlings S., Hill G. J., Mitchell E., Dunlop J. S., Wold M., 2004, *MNRAS*, 351, 347
- Magliocchetti M., Maddox S. J., Lahav O., Wall J. V., 1999, *MNRAS*, 306, 943
- Magliocchetti M. et al., 2004, *MNRAS*, 350, 1485
- Matarrese S., Coles P., Lucchin F., Moscardini L., 1997, *MNRAS*, 286, 115
- Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347
- Nilsson K. K., Meisenheimer K., 2009, *New Astron. Rev.*, 53, 37
- Norberg P. et al., 2002, *MNRAS*, 332, 827
- Norris R. P. et al., 2011, *Publ. Astron. Soc. Aust.*, 28, 215
- Overzier R. A., Röttgering H. J. A., Rengelink R. B., Wilman R. J., 2003, *A&A*, 405, 53
- Passmoor S., Cress C., Faltenbacher A., Johnston R., Smith M., Ratsimbazafy A., Hoyle B., 2013, *MNRAS*, 429, 2183
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
- Percival W. J. et al., 2007, *ApJ*, 657, 645
- Percival W. J. et al., 2010, *MNRAS*, 401, 2148
- Raccanelli A., Bonaldi A., Negrello M., Matarrese S., Tormen G., de Zotti G., 2008, *MNRAS*, 386, 2161
- Raccanelli A. et al., 2012, *MNRAS*, 424, 801
- Rengelink R. et al., 1998, in Bremer M. N., Jackson N., Perez-Fournon I., eds, *Astrophysics and Space Science Library*, Vol. 226, *Observational Cosmology with the New Radio Surveys*. Kluwer, Dordrecht, p. 143
- Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
- Roche N. D., Almaini O., Dunlop J., Ivison R. J., Willott C. J., 2002, *MNRAS*, 337, 1282
- Ross N. P. et al., 2009, *ApJ*, 697, 1634
- Saunders W., Rowan-Robinson M., Lawrence A., 1992, *MNRAS*, 258, 134
- Scranton R. et al., 2005, *ApJ*, 633, 589
- Seljak U., Warren M. S., 2004, *MNRAS*, 355, 129
- Simpson C. et al., 2012, *MNRAS*, 421, 3060
- Smith D. J. B. et al., 2011, *MNRAS*, 416, 857
- Smolčić V. et al., 2009, *ApJ*, 696, 24
- Sullivan M., Hopkins A. M., Afonso J., Georgakakis A., Chan B., Cram L. E., Mobasher B., Almeida C., 2004, *ApJS*, 155, 1
- Sutherland W., Saunders W., 1992, *MNRAS*, 259, 413
- Wake D. A., Croom S. M., Sadler E. M., Johnston H. M., 2008, *MNRAS*, 391, 1674
- Wang L. et al., 2011, *MNRAS*, 414, 596
- White R. L., Becker R. H., Helfand D. J., Gregg M. D., 1997, *ApJ*, 475, 479
- Willmer C. N. A., da Costa L. N., Pellegrini P. S., 1998, *AJ*, 115, 869
- Willott C. J., Rawlings S., Blundell K. M., 2001, *MNRAS*, 324, 1
- Willott C. J., McLure R. J., Jarvis M. J., Rawlings S., Hill G. J., Mitchell E., Dunlop J. S., 2003, *New Astron. Rev.*, 47, 187
- Wilman R. J. et al., 2008, *MNRAS*, 388, 1335
- Wilman R. J., Jarvis M. J., Mauch T., Rawlings S., Hickey S., 2010, *MNRAS*, 405, 447
- Yang A., Saslaw W. C., 2011, *ApJ*, 729, 123
- York D. G. et al., 2000, *AJ*, 120, 1579

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.